DOCUMENT RESUME

ED 050 741 LI 002 800

AUTHOR Herling, John P.; Lazorick, Gerald J.
TITLE Proposal for a Current Awareness Service for the Literature of Library and Information Science.
INSTITUTION State Univ. of New York, Albany.
PUB DATE 68
NOTE 27p.; Paper prepared for the Conference on the Bibliographic Control of Library Science Literature, State University of New York at Albany, April 19-20, 1968

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Communication (Thought Transfer), Computers, Conferences, Government Publications, Information Needs, *Information Science, *Library Science, *Permuted Indexes, Scholarly Journals, Vocabulary Development
IDENTIFIERS Keyword Out of Context, *Library Science Literature, Machine Readable Data

ABSTRACT

To meet the need for quick access to the literature, and to obtain more effective communication among librarians, information scientists, mass communication specialists, operations analysts, and behavioral scientists, the authors propose a KWOC (Key-Word Out-of-Context) Index using existing machine-readable bibliographical data, such as supplied by the Institute of Scientific Information and Pandex, supplemented by in-house conversion of core journals and U.S. Government Research Reports. Some 340,000 items (annually) would be matched against a set of authors' names, journal titles, and index terms or groups of terms, the latter weighted to eliminate false drops. Preliminary computer output would be manually screened, then matched against an open vocabulary to provide the output in the form of a monthly printed bulletin arranged by subject, with an author index. (Related documents are LI 002 796 - 002 799, and LI 002 801 - 002 807). (Author)

ED050741

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

PROPOSAL FOR A CURRENT AWARENESS SERVICE FOR THE LITERATURE OF LIBRARY AND INFORMATION SCIENCE

by

John P. Herling
Assistant Director of Libraries
State University of New York at Buffalo

and

Gerald J. Lazorick, Director
Technical Information Dissemination Bureau
State University of New York at Buffalo

LL 002 800

CONFERENCE ON THE BIBLIOGRAPHIC CONTROL OF LIBRARY SCIENCE LITERATURE

State University of New York at Albany
April 19-20, 1968

2

ED050741

INTRODUCTION

It is an ironic fact that control of the literature of bibliographical control - the literature of librarianship - the literature of information science and technology - is inadequate. Librarians who frequently deplore the inadequacies of abstracting and indexing services in most disciplines, have themselves produced no prompt, comprehensive services for their own discipline. Although there seems to be a growing awareness of the problem, it is only limited, for one does not find improved methods of control of library literature among the forty four major goals for action of the American Library Association in 1967 (1), nor in the recent survey in one of the leading library journals of trends in bibliography in twenty five areas of knowledge, is there a paper devoted to trends in the bibliography of librarianship (2).

There are important and useful bibliographic services in library and information science. H. Allan Whately (3) identified twenty one. We do not denigrate such services as Library Literature and Library Science Abstracts. More recent additions to the bibliography of this field Documentation Abstracts and the Annual Review of Information Science and Technology are valuable contributions to the control of the literature. However, most of us would agree that none of the currently published abstracting and indexing services is doing a good job in respect to coverage and promptness. Alan Gilchrist, who checked a bibliography on the evaluation of information systems against six services, found that the best coverage (68%) was achieved

by American Documentation (in its "Literature Notes," now Documentation Abstracts) and by Library Literature (58%). The time lag between publication of originals and publication of abstracts averaged 213 days in Library Literature and 281 days in American Documentation (4). Whatley (5) comments on the failure of the services to cover the report literature and cites the editors' estimates of a three to six month time lag. We do not propose to solve all the problems of bibliographical control of the literature of library and information science and technology, nor do we believe there is "one best way". We agree with Phyllis Richmond (6) that, "Rather than trying to achieve the impossibility of being all things to all men, which is again giving at least lip-service to the principle of "the one best way", it is more prudent to have several systems with differing objectives and differing levels of complexity or detail, as well as varying design, composition and operation, and to use them in conjunction with each other, both directly in some cases and indirectly in others - all in a shotgun rather than a rifle approach."

The system we propose is a computer-based current awareness service for the literature of library and information science and technology. No one can deny that we are in a period of explosive developments in librarianship when quick access to the latest results of experimentation and research and to the latest information on new methods and new equipment for the acquisition, recording and transfer of information is of paramount importance.

The design of our system and our approach to current awareness will be discussed later. We should like to emphasize, here, our conviction of the need for such a service. Whatley's survey of the purposes for which librarians and library school staff and students used present abstracting and indexing services indicated that the primary purpose was current awareness (7). Although this survey employed only the questionnaire technique and was based on a relatively small sample, it is probably the most extensive that has been made in the field of library studies. Previously published studies of librarians as users of library literature are scarcely worth citing. But this is not unique to librarianship for user studies in general (most of these are in the scientific information field) are of questionable value (8). Saul Herner (9) at the conclusion of an analysis of the crudeness of user studies makes the point that though there are signs of convergence, there are still two different groups of people working in the information field: [1] the working librarians and information specialists and [2] the mass communication specialists, operation analysts, and behavioral scientists. Only when we have obtained more effective communication within the field, he says, can we realistically aspire to understand and interpret the communication and information gathering habits of its clientele. Current awareness tools, he adds, are one of the means to obtain more effective communication within the field.

THE APPROACH TO CURRENT AWARENESS

We will design the system to keep a library or information scientist abreast of what is happening in his field of interest,

and as a service to organizations trying to remain in the forefront of this rapidly expanding field. The system is based on the premise that information is most valuable if it is supplied and digested as soon as possible after publication. In this way, the scientist or practitioner can more quickly transfer information into new, improved or expanded systems and services.

To accomplish the current awareness objective, the system supplying the information must be fast. It should require as little of the user's time as possible. It should be economical - the cost to benefit ratio must be as low as possible. It should have high precision - the information the man receives should be pertinent to his interests. He should have to scan as little irrelevant information as possible. Our system will computer scan each of the titles and authors of approximately 341,500 items currently appearing in more than 2,000 scientific and technical journals, the government research report literature available from the Clearinghouse for Federal Scientific and Technical Information and patent information which may soon be available in machine readable form. The data bank will also include information about new scientific and technical books published in the United States. The system identifies published items which should be of interest to the individual. By pointing a man toward information which he does not regularly see, it supplements, but is not a substitute for, the man's own attempts to keep abreast of the expanding technologies.

The author-title information is indexed, using a modified version of Luhn's derivative indexing methods. Thus the advantages of

this type of permuted indexing and the outgrowth of his method - KWIC (Keyword in Context) and KWOC (Keyword out of Context) - are inherent in our method of computer scanning. Tukey (10) stated that the permuted index was invented as a means of adequately controlling (essentially of indexing) the literature without extensive intellectual effort, and thus eliminating indexing delays. Cleverdon (11) commented that the great merit of this particular method . . . is that it enables information concerning new articles to be made available much more quickly than if there were the inevitable delays of human abstracting and indexing. Janaske (12) indicates that - though there are disadvantages which he points out - perhaps the greatest advantage is the timeliness and the speed with which permuted title indexes can be prepared. Skaggs and Spangler (13) tested a permuted indexing system and stated, "The most obvious advantage of permuted indexing by computer is speed. In a test of one permuted indexing system, input of 3,000 cards containing titles and running text produced a permuted significant word index of 12,190 index entry lines, with approximately 85 minutes of computer time required for the permuting and sort operation." Kennedy (14) states, "that the use of the author's own terms . . . the alive currency of new ideas . . . rather than the considered reshapings of the indexing system may often be of advantage. The automatic generation as index entries of all the separate words in multi-term concepts is definitely so. Access is direct, under any one of the component terms, in the unrestricted manner of uniterm indexing and context minimizes

false drops; the author has supplied the term coordination." Stevens (15) points out that, in general, even among enthusiasts of KWIC, there is more agreement as to the values of the technique as a device for current awareness scanning and as a dissemination index than its use for more extensive searching. She indicates that, in fact, it was primarily as a dissemination index that Luhn first proposed the KWIC technique. Ruhl (16) found that between 50 and 90 percent of author prepared titles (the variation depending on subject field and other circumstances) did fully reflect the index terms assigned to these documents by human indexers. Stevens (17) also states that, "Lane and White and Walsh have also made studies directly related to the question of KWIC index effectiveness." The latter two investigators report only 52 percent retrieval effectiveness for a permuted title index to the Abstracts of Computer Literature - 1962 which they attribute to the changing terminology in the still new field of computer technology. Lane (18) made counts of titles that would be "acceptable" and those that would not for a KWIC index for 50 titles drawn from each of 10 published indexes. He concluded that, if there were judicious pre-editing, technical articles in the technical subject indexes could be adequately covered, and papers in the field of law, business and humanities somewhat less satisfactorily so, but that for material indexed in the Reader's Guide to Periodical Literature, the KWIC technique would fail 58 percent of the time. Montgomery and Swanson (19) have studied the adequacy of "machine-like indexing by people." Montgomery and Swanson took as their test corpus the September, 1960 issue of Index Medicus and found that for 4,770

items, 85.8 percent contained either the word itself or a synonym for the subject heading assigned, slightly over 11 percent did not, and in the remaining cases the investigator could not clearly decide. The National Science Foundation's CR&D Report No. 11 (20) cites a study of the Index to Legal Periodicals. The report states, "Interpretation of data revealed, among other things, that 64.4 percent of the title entries contained as keywords one or more of the ILP subject heading words under which they were indexed and 25.1 percent contained logical equivalents. The remaining 10.5 percent of the title entries had nondescriptive titles."

Many individuals compare computer generated indexes with human indexing and comment about the superior quality of human indexing. Researchers have studied this and also the speed of human indexers. Cleverdon (21) states that few reliable figures have been given for current practices, although a particularly high figure is the 1 1/2 hours average quoted for indexing reports for the catalogue of aerodynamic data prepared by the National Fluchtvaart Laboratories in Holland. Further, he points out that it appears from personal discussions that 20 minutes per report for a general collection of technical reports is the top limit, and this has been taken as the maximum indexing time to be used in the project.

Black (22) states, "that there have been enough experiments to indicate that there is no consistency, or very little, between one indexing performance by a given individual and another indexing performance at a later date, by the same individual.

The same inconsistency has been discovered among different individuals all indexing the same documents." Thus, there is neither inter-indexer nor intra-indexer consistency in a system that depends on human performance. Stevens (23) states, "There can be little doubt that quality and consistency of most human indexing, practically available today, is not good. Much of it, because of time and other pressures, is either directly a word extraction process, or it is inconsistent in assignment of many relevant descriptors and subject category labels. On the other hand, today's indexing, whether accomplished by man or machine, is probably no better or no worse than any other classificatory or indexing procedures. The only excuse, therefore, for choice between man and machine is the cost/benefit ratio which is related on the one hand to specific operational considerations and on the other to the question of whether or not various users would agree with the machine as much as they agree with each other." Black (24) writes, "that it has been estimated that the efficiency of KWIC indexing is about 76 percent compared with about 82 percent for the conventional indexing or classification."

Perhaps the most significant comment of all was made by Wyllys (25):

> "It is well-known that the current methods of producing, through human efforts, condensed representations of documents are already hopelessly inadequate to cope with the present volume of scientific and technical literature. Many papers are never indexed or abstracted

at all, and even in the cases of those that are indexed or abstracted, the indexes and abstracts do not become available until six months to two years after the publication of the paper."

THE INPUT TO THE SYSTEM

As a data base we will use machine readable tapes, such as those presently available from Pandex and the Institute of Scientific Information, which cover over 2,000 scientific and engineering journals including the following important titles in library and information science:

Aslib Proceedings

American Documentation

Bulletin of the Medical Library Association

Information Storage and Retrieval

Journal of Chemical Documentation

Journal of Documentation

Library Resources and Technical Services

Methods of Information in Medicine

Nachrichten fur Dokumentation

Technical Information Center Administration

Other journals of related interest, available on these tapes are:

American Behavioral Scientist

Association for Computing Machinery, Communications

Computer Bulletin

Computer Design

Computing

Data Processing

Datamation

IBM Journal of Research and Development

IEEE Spectrum

IEEE Transactions

Operations Research

Science

To supplement these data, machine readable information concerning all current technical reports available from the Clearinghouse for Federal Scientific and Technical Information as found in the publication, U.S. Government Research and Development Reports is already available at the Technical Information Dissemination Bureau. We will ourselves convert to machine readable form the bibliographic information for all substantive articles appearing in the following journals:

ALA Bulletin

Aktulne Problemy Informacji i Dokumentacji (Warsaw)

Annals of Library Science and Documentation

Association Internationale łes Documentalistes Bulletin

Atlantic Provinces Library Association Bulletin

Bibliography, Documentation Terminology

Bulletin des Bibliotheques de France

Cahiers de la Documentation

Canadian Library

College and Research Libraries

D. C. Libraries

Dokumentation

Drexel Library Quarterly

Herald of Library Science

Illinois Libraries

Indexer

Indian Association of Special Libraries and Information Centres, Bulletin

Journal of Education for Librarianship

Journal of Library History

Knihovnik (Prague)

Law Library Journal

Library Association Record

Library Journal

Library Quarterly

Library Review

Library Science with a Slant to Documentation

Library Trends

Library World

Libri

Metodica A Tehnika Informaci (Prague)

Music Library Association Notes

National Micro-News

Nauchno-Technicheskaya Informatsiya

New Zealand Libraries

North Carolina Libraries

Ohio Library Association Bulletin

Ontario Library Review

PNLA Quarterly

Pennsylvania Library Association Bulletin

Problemy Peredachi Informatsii

Scientific Information Notes

Southeastern Libraries

South African Libraries

Special Libraries

Technicheski Biblioteki SSSR

Tennika Knihovna (Prague)

Tidskrift för Dokumentation (Stockholm)

Tijdschrift voor Efficiente e Documentatie

Tudomanyos es Muszaki Tajekoztatas

UNESCO Bulletin for Libraries

Verband der Bibliotheken des Landes Nordrhein - Westfalen, Mitteilungsblatt

Wilson Library Bulletin

Zeitschrift fur Bibliothekswesen

In addition, by means of a computer-produced mailing, we will solicit from individuals and organizations engaged in research and development in library and information science, reports of research in progress, new technical reports, and preprints of papers, thus augmenting the informal channels of communication that play an important part in the transfer of information in all fields (26). We shall also publish advertisements of the service in selected journals, in which we will request reports and papers to be submitted for indexing. 6,000 new book titles (yearly) are also already available in machine readable form.

The information for each item in the bank consists of, - at least - authors, title, journal abbreviation, volume, page numbers, and a unique identifying number. Non descriptive titles have subtitles appended to them (if a subtitle is used) or the first portion of an abstract (up to 174 characters) if an abstract is available. Many of the journal article titles are enriched with index terms. All report titles are enriched with index terms. Report information consists of authors, title, index terms, report number (AD, PB, etc.) and corporate author. Contract numbers are also given. Each book title is enriched with about ten index terms.

Thus, the base we will search annually will total an estimated 341,500 articles, documents, and book titles comprising the following:

1)	From journals available on tape	300,000
2)	Clearinghouse Reports Converted to tape	30,000
3)	Library journals converted to tape	4,500
4)	Contributed papers and reports	1,000
5)	New book titles	6,000
		341,500

The estimated annual pertinent items of information totals 7225 as follows:

From 1, above	1,200
From 2, above	500
From 3, above	4,500
From 4, above	1,000
From 5, above	25
	7,225

We would, in addition, investigate other sources of machine-readable bibliographical data produced by documentation centers, particularly in Western and Eastern Europe, with a view to cooperative exchange.

THE OUTPUT OF THE SYSTEM

The output of the proposed system will be the following monthly computer produced indexes:

1) A list of articles and reports arranged alphabetically by all authors.
2) A keyword out of context (KWOC) index.

A sample format from another current awareness system which is now operational at the Technical Information Dissemination Bureau is shown in Figure 1. The lists will be reduced to 8 1/2" x 11" size and printed using an economically appropriate process. The exact format of the computer output in the keyword out of context list will probably be ordered by first author in order to facilitate searching the index using groups of keywords.

It is estimated that the monthly output will contain approximately six hundred entries. These will appear an average of two times in the author list and five times in the KWOC list. Thus, we anticipate that the monthly list will be printed on approximately 200 pages (100 leaves).

Figure 1

Sample KWOC Index

ENERGY
POSSIBLE EXPLANATION OF PROTECTIVE ACTION OF POLYANIONS AND CYSTEINE IN RADIOLYSIS BASED ON MECHANISM OF ENERGY TRANSFER
YERMOLAY.VK CHEREPAN.VY BERDYSHE.GD MOLIN YN
SEL. NO.-0078272 SOURCE-BIOPHYSIC R V- 11 P- 924 YR-66

ENZYMES
PHYSICOCHEMICAL INVESTIGATION OF MECHANISM OF INACTIVATION OF ENZYMES .3. THERMODYNAMIC ANALYSIS OF STRUCTURAL-CHEMICAL CHANGES IN PROTEINS ON RADIATION INACTIVAC
EIDUS LK
SEL. NO.-0078271 SOURCE-BIOPHYSIC R V- 11 P- 684 YR-66

ESR
ESR SPECTRA OF A SODIUM BORATE GLASS IRRADIATED BY ULTRA-VIOLET RAY
KISHII T OOKA K
SEL. NO.-0081141 SOURCE-J PHYS JAP V- 23 P- 659 YR-67

ETCHED
A COMPARISON BETWEEN ETCHED FISSION TRACK DENSITIES ON INTERNAL AND EXTERNAL GLASS SURFACES AFTER NEUTRON IRRADIATION
EDWARDS J
SEL. NO.-0081136 SOURCE-GEOPHYS J R V- 13 P- 541 YR-67

EUROPES
EUROPES LATEST IN PROCESSING - AMONG ADVANCES - FILLET IRRADIATION CONTINUOUS SPIRAL FREEZING MICROWAVE STERILIZATION FLUIDIZED SOUP MIXING AUTOMATED SUGAR PR P WOOLLEN AH
SEL. NO.-0079386 SOURCE-FOOD ENG V- 39 P- 78 YR-67

EVALUATION
ADENOCARCINOMA OF ENDOMETRIUM - EVALUATION OF PREOPERATIVE IRRADIATION AND FACTORS INFLUENCING PROGNOSIS
WADE ME KOHORN EI MORRIS JC
SEL. NO.-0079379 SOURCE-AM J OBST G V- 99 P- 869 YR-67

EXAMINES
FOOD IRRADIATION R&D CONTAINUES - INTEREST IN RADIATION PRESERVATION OF FOOD WIDENS AS RESEARCH EXAMINES FEASIBILITY OF SPECIFIC PROCESSES . PILOT-SCALE ACTICTI
VARNER DM
SEL. NO.-0079387 SOURCE-FOOD ENG V- 39 P- 93 YR-67

EXCHANGE
CATALYTIC DEUTERIUM EXCHANGE REACTIONS WITH ORGANICS .31. ACCELERATIVE EFFECTS OF IONIZING RADIATION AND ULTRAVIOLET LIGHT ON SELF-ACTIVATED PLATINUM OXIDE E
GARNETT JL RAINIS A WONG PK
SEL. NO.-0081133 SOURCE-AUST J CHEM V- 20 P-2415 YR-67

EXCITATION
LOW-ENERGY ELECTRON SCATTERING BY ATOMIC HYDROGEN .3. COMPARISON OF THEORY AND EXPERIMENT FOR ELECTRON-INDUCED EXCITATION OF HYDROGEN TO N#2 LEVEL
BURKE PG TAYLOR AJ ORMONDE S
SEL. NO.-0081146 SOURCE-P PHYS SOC V- 92 P- 345 YR-67

EXPERIMENT
LOW-ENERGY ELECTRON SCATTERING BY ATOMIC HYDROGEN .3. COMPARISON OF THEORY AND EXPERIMENT FOR ELECTRON-INDUCED EXCITATION OF HYDROGEN TO N#2 LEVEL
BURKE PG TAYLOR AJ ORMONDE S
SEL. NO.-0081146 SOURCE-P PHYS SOC V- 92 P- 345 YR-67

EXPLANATION
POSSIBLE EXPLANATION OF PROTECTIVE ACTION OF POLYANIONS AND CYSTEINE IN RADIOLYSIS BASED ON MECHANISM OF ENERGY TRANSFER
YERMOLAY.VK CHEREPAN.VY BERDYSHE.GD MOLIN YN
SEL. NO.-0078272 SOURCE-BIOPHYSIC R V- 11 P- 924 YR-66

EXTERNAL
A COMPARISON BETWEEN ETCHED FISSION TRACK DENSITIES ON INTERNAL AND EXTERNAL GLASS SURFACES AFTER NEUTRON IRRADIATION
EDWARDS J
SEL. NO.-0081136 SOURCE-GEOPHYS J R V- 13 P- 541 YR-67

FACTORS
ADENOCARCINOMA OF ENDOMETRIUM - EVALUATION OF PREOPERATIVE IRRADIATION AND FACTORS INFLUENCING PROGNOSIS
WADE ME KOHORN EI MORRIS JC
SEL. NO.-0079379 SOURCE-AM J OBST G V- 99 P- 869 YR-67

FADING
FADING IN THERMOLUMINESCENT LITHIUM FLUORIDE AFTER IRRADIATION WITH GAMMA-RAY DOSES UP TO 1000 RAD
WEBB GAM
SEL. NO.-0081134 SOURCE-BR J A PHYS V- 18 P-1567 YR-67

FAST
MASS AND ENERGY DISTRIBUTIONS OF FRAGMENTS FROM U233 FISSION INDUCED BY THERMAL AND FAST NEUTRONS
SENCHENK.VI TARASKO MZ MIKHAILO.VB KUZMINOV BD
SEL. NO.-0081153 SOURCE-SOV J NUC R V- 5 P- 362 YR-67

FEASIBILITY
FOOD IRRADIATION R&D CONTAINUES - INTEREST IN RADIATION PRESERVATION OF FOOD WIDENS AS RESEARCH EXAMINES FEASIBILITY OF SPECIFIC PROCESSES . PILOT-SCALE ACTICTI
VARNER DM
SEL. NO.-0079387 SOURCE-FOOD ENG V- 39 P- 93 YR-67

FERRITES-
EFFECT OF GAMMA AND NEUTRON IRRADIATION ON FERROMAGNETIC RESONANCE IN FERRITES-
LEBED BM MUKHA LY MOSEL VI JLMANIS JA
SEL. NO.-0081156 SOURCE-SOV PH SS R V- 9 P-1342 YR-67

FERROMAGNETIC
EFFECT OF GAMMA AND NEUTRON IRRADIATION ON FERROMAGNETIC RESONANCE IN FERRITES-
LEBED BM MUKHA LY MOSEL VI JLMANIS JA
SEL. NO.-0081156 SOURCE-SOV PH SS R V- 9 P-1342 YR-67

FILLET
EUROPES LATEST IN PROCESSING - AMONG ADVANCES - FILLET IRRADIATION CONTINUOUS SPIRAL FREEZING MICROWAVE STERILIZATION FLUIDIZED SOUP MIXING AUTOMATED SUGAR PR P WOOLLEN AH
SEL. NO.-0079386 SOURCE-FOOD ENG V- 39 P- 78 YR-67

FILM
MAGNETIC PROPERTIES OF NEUTRON-IRRADIATED IRON-NICKEL THIN FILM ALLOY
TANAGAWA N SUGIURA Y HARASAWA S
SEL. NO.-0081142 SOURCE-J PHYS JAP V- 23 P- 908 YR-67

FINISHING
IRRADIATION FINISHING APPLICATIONS
STANNETT V
SEL. NO.-0078266 SOURCE-AM DYE REP V- 55 P- 32 YR-67

FISSION
A COMPARISON BETWEEN ETCHED FISSION TRACK DENSITIES ON INTERNAL AND EXTERNAL GLASS SURFACES AFTER NEUTRON IRRADIATION
EDWARDS J
SEL. NO.-0081136 SOURCE-GEOPHYS J R V- 13 P- 541 YR-67

MASS AND ENERGY DISTRIBUTIONS OF FRAGMENTS FROM U233 FISSION INDUCED BY THERMAL AND FAST NEUTRONS
SENCHENK.VI TARASKO MZ MIKHAILO.VB KUZMINOV BD
SEL. NO.-0081153 SOURCE-SOV J NUC R V- 5 P- 362 YR-67

CHANNEL ANALYSIS OF NEUTRON-INDUCED FISSION OF U234-
VOROTNIK.PE
SEL. NO.-0081154 SOURCE-SOV J NUC R V- 5 P- 415 YR-67

FLUIDIZED
EUROPES LATEST IN PROCESSING - AMONG ADVANCES - FILLET IRRADIATION CONTINUOUS SPIRAL FREEZING MICROWAVE STERILIZATION FLUIDIZED SOUP MIXING AUTOMATED SUGAR PR P WOOLLEN AH
SEL. NO.-0079386 SOURCE-FOOD ENG V- 39 P- 78 YR-67

FLUORIDE
FADING IN THERMOLUMINESCENT LITHIUM FLUORIDE AFTER IRRADIATION WITH GAMMA-RAY DOSES UP TO 1000 RAD
WEBB GAM
SEL. NO.-0081134 SOURCE-BR J A PHYS V- 18 P-1567 YR-67

FOOD
FOOD IRRADIATION R&D CONTAINUES - INTEREST IN RADIATION PRESERVATION OF FOOD WIDENS AS RESEARCH EXAMINES FEASIBILITY OF SPECIFIC PROCESSES . PILOT-SCALE ACTICTI
VARNER DM
SEL. NO.-0079387 SOURCE-FOOD ENG V- 39 P- 93 YR-67

FORMATION
PHOTOCHEMICAL STUDIES .9. FORMATION OF BENZ@)@-1,3-OXAZEPINES IN PHOTOLYSIS OF QUINOLINE-N-OXIDES IN SOLUTION
BUCHARDT O JENSEN B LARSEN IK
SEL. NO.-0081126 SOURCE-ACT CHEM SC V- 21 P-1841 YR-67

RADIATION CHEMISTRY OF GASEOUS AMMONIA .2. HYDRAZINE FORMATION
JONES FT SWORSKI TJ WILLIAMS JM
SEL. NO.-0081161 SOURCE-T FARAD SOC V- 63 P-2426 YR-67

FRAGMENTS
MASS AND ENERGY DISTRIBUTIONS OF FRAGMENTS FROM U233 FISSION INDUCED BY THERMAL AND FAST NEUTRONS
SENCHENK.VI TARASKO MZ MIKHAILO.VB KUZMINOV BD
SEL. NO.-0081153 SOURCE-SOV J NUC R V- 5 P- 362 YR-67

FREEZING
EUROPES LATEST IN PROCESSING - AMONG ADVANCES - FILLET IRRADIATION CONTINUOUS SPIRAL FREEZING MICROWAVE STERILIZATION FLUIDIZED SOUP MIXING AUTOMATED SUGAR PR P WOOLLEN AH
SEL. NO.-0079386 SOURCE-FOOD ENG V- 39 P- 78 YR-67

FRYER
LOW-DOSE IONIZING IRRADIATION OF TRAY-PACKED CUT-UP FRYER CHICKENS
MERCURI AJ KOTULA AW SANDERS DH
SEL. NO.-0081135 SOURCE-FOOD TECHN V- 21 P-1509 YR-67

FUEL
INVESTIGATION ON HEAT TRANSFER BETWEEN FUEL AND CLADDING AND ON INTERACTIONS BETWEEN FUEL AND CLADDING DURING SHORT-DURATION IRRADIATIONS UNDER BOILING WATER R
SCHWARZW.R
SEL. NO.-0078267 SOURCE-ATOMKERNENE V- 12 P- 343 YR-67

THE SYSTEM

The system will be composed of two phases - a retrieval phase and a dissemination phase. The edited output from the Retrieval Phase will form the input for the Dissemination Phase. Each phase will now be described in turn.

Retrieval Phase

In this phase, a set of authors names, journal titles and index terms (or groups of terms) from an open vocabulary will be matched monthly with machine readable information added to the data bank during the past month. The search will be performed using word stems. In this manner, a word stem such as "Librar" will match Library, Libraries, Librarian, Librarians, Librarianship, etc. Each term or term group will be given a weight between 0 and 1.0. A match will always occur when an authors name matches or when a journal title matches. (An eleven digit alphameric code is used for the journal title). Terms or term groups will be of two types - primary or secondary. Primary terms which match will always generate a "hit". Secondary terms will not generate a "hit" unless they occur in combination with another secondary term. The combining of only secondary terms can generate a "false drop." This can be controlled by judicious term weighting. The index terms or term groups can be assigned weights between 0.0 and 1.0. Each term weight may be considered to be a probability. That is, an item containing the term which has weight, say .3, has probability .3 of being a pertinent item. The weights of all terms or term groups are

combined using the equation for the probability of the union of two events. Thus if

T1 = Weight of Term 1

T2 = Weight of Term 2

The weight for the match is

$$T = T1 + T2 - (T1)\ (T2)$$

The value of T is passed against a threshold value and

If $T \geqslant TV$ Match

$T \lessdot TV$ No Match

Some examples of terms in the vocabulary are:

Index

Cost

Effectiveness

Library

Some examples of term groups in the vocabulary are:

Information Retrieval

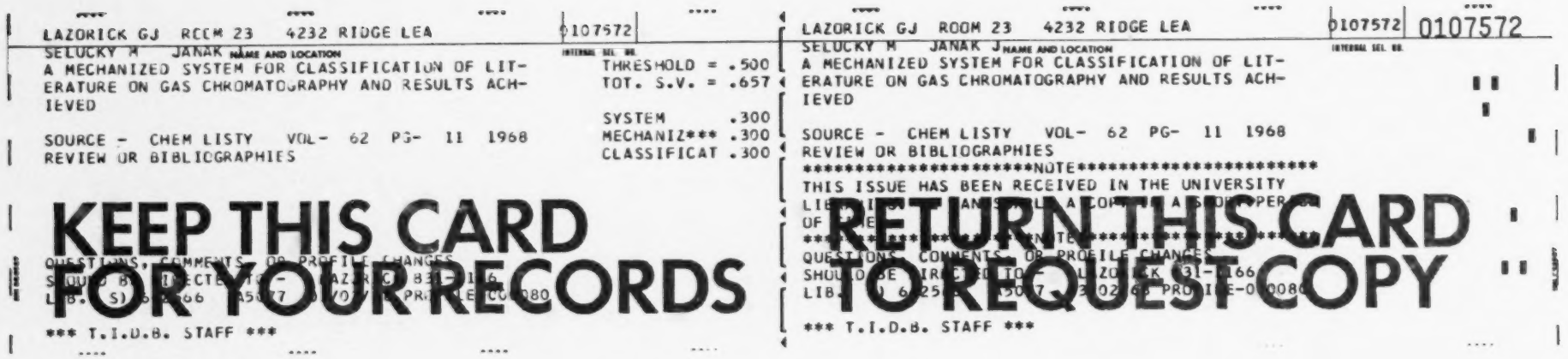
Information Storage

Information Dissemination

The exact term weights which are assigned depend on the search strategy which is used. The search strategy selected is not independent of the system. Using the terms above, let's say with a threshold value of .5, we could weight all 3 of the latter term groups .5. Each then would generate a hit. Alternately, one could weight "Information" at say .3 and each of the terms "Retrieval", "Storage" and "Dissemination" at .3. Either method would retrieve an article which had the term group "Information Retrieval" in the title. The one strategy however

would result in a total significance value of .5 whereas the other would result in a significance value of .51 (.3 + .3 - .09). However, the one method may result in considerably more computer time and "noise" than the other because in the former case the word group contains the term "storage" in a group whereas in the latter case the term "storage" occurs alone. Even though the term weight is low and would not by itself generate a "hit", it will generate a "hit" when combined with another low weight term and thus cause system noise. One is then faced with deriving a lengthy open vocabulary of highly specific term groups or a shorter vocabulary of individual terms which will generate more system noise. In the former case, computer time increases as the number of term groups increases. In the latter case, computer time increases because the number of hits increases. Further, the former case will result in higher precision and lower recall than the latter. The authors favor higher recall and lower precision since it is proposed that all hits will be humanly edited prior to input into the dissemination phase. Information about items whose total significance value is greater than the threshold value will be output on a two part card (Figure 2). Also, as output, there will be a machine readable record of the "hit" on magnetic tapes. These will form the input for the dissemination phase.

An editor will then scan each of the cards and determine which will be of interest or of peripheral interest to the field. He will remove the right hand portion of the pertinent cards.



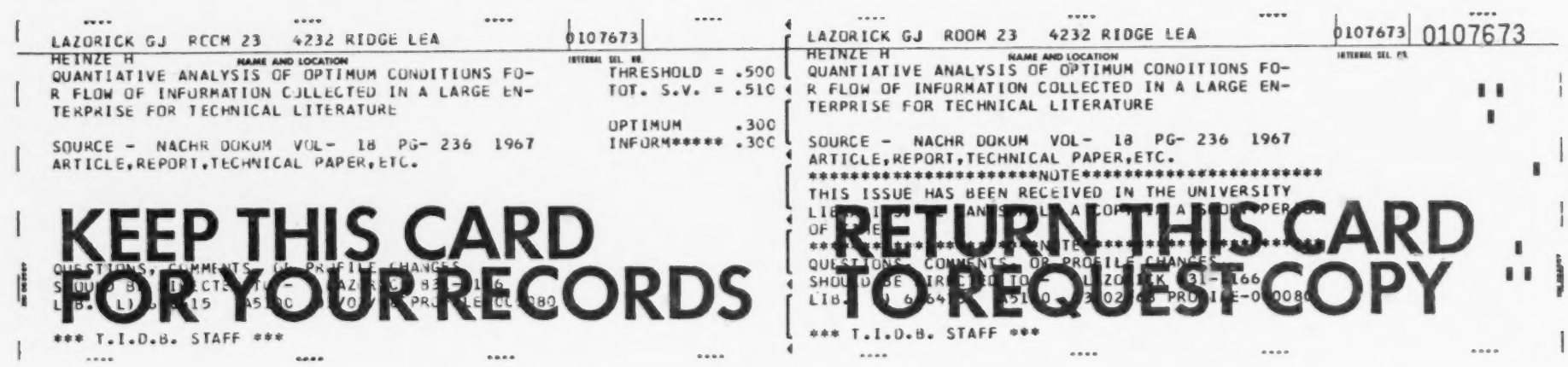

LAZORICK GJ ROOM 23 4232 RIDGE LEA
0107673
REINZE H NAME AND LOCATION
INTERNAL SEL. NO.
QUANTIATIVE ANALYSIS OF OPTIMUM CONDITIONS FO-
R FLOW OF INFORMATION COLLECTED IN A LARGE EN-
TERPRISE FOR TECHNICAL LITERATURE
SOURCE - NACHR DOKUM VOL- 18 PG- 236 1967
ARTICLE,REPORT,TECHNICAL PAPER,ETC.
THRESHOLD = .500
TOT. S.V. = .510
OPTIMUM .300
INFORM***** .300
KEEP THIS CARD
FOR YOUR RECORDS
RETURN THIS CARD
TO REQUEST COPY
QUESTIONS, COMMENTS, OR PROFILE CHANGES
*** T.I.D.B. STAFF ***
0107673

Figure 2
Sample Cards

These selected "hits" (each right hand card is prepunched with an identification number) together with the tape output form the input to the dissemination system.

We will use a dynamic open vocabulary system in the sense that terms and term groups will continuously be added. The terms will be added when they co-occur (with high frequency) in "hit" items with terms already in the open vocabulary. The terms will be selected by humanly scanning a computer produced word frequency ordered list of terms in "hit" items which are not already in the vocabulary. In this way, new terminology will always be added as it is introduced in the field. The open vocabulary will grow and the word frequency list will diminish. Prior to running this retrieval system, all data bank items will be passed against a stop list of high frequency common terms. The remaining terms will be alphabetically sorted as will the open vocabulary index terms. A term by term match will then take place. Common nondescriptive terms from the term frequency list previously described will be entered into the stop list. Thus, the stop list will grow in size. (A page from a sample stop list is shown in Figure 3).

The Dissemination Phase

The Dissemination Phase will use as input the selected cards and magnetic tape output from the retrieval phase as well as the list of terms which will be used for the KWOC. The items selected by the editor will not be indexed by all significant words in the title. Each title will be passed against a broad

Figure 3

Stop List Page

A	CONS	FURTHER	NOTE
ABOUT	CONTAINING	FUTURE	NOW
ABREAST	DANS	GAP	OCCASION
ACCOMPLISHI	DAS	GENERAL	OF
ACCORDING	DE	GET	OFF
ADDS	DEM	GETS	ON
ADVANCES	DEN	GIVEN	ONLY
AFTER	DER	GOODBY	OR
AGAINST	DES	GREAT	OTHER
ALL	DETERMINATI	HAS	OU
ALSO	DEVELOPED	HAVE	OUT
AM	DI	HELP	OVER
AMENABLE	DID	HERE	PREVIOUS
AMONG	DIE	HOW	PRIMARILY
AN	DIFFERENT	I	PROS
AND	DISCUSSION	II	PUSH
ANNUAL	DO	IM	PUT
ANON	DOES	IN	PUTTING
ANOTHER	DU	INTERESTING	RATIONALE
ANY	DUE	INTO	RELATED
APPEAR	DUNE	INTRODUCTIO	REMARD
ARE	DURCH	IS	RESULT
ARISING	DURING	IT	RESULTS
AS	EASES	ITS	REVIEW
AT	EASIER	IV	ROSY
AU	EASY	IX	SAY
AUF	EFFET	KEEPING	SEE
AUS	EFFETS	KNOWABLE	SEEN
AUX	EIN	LA	SEVERAL
AVEC	EINE	LE	SHOULD
BE	EINEM	LES	SO
BEI	EINER	LETUDE	SOME
BEING	EINES	LIVELY	SOON
BEST	EN	LOOKING	SUGGESTED
BETTER	ENTRE	LOOKS	SUR
BETWEEN	EPILOGUE	MADE	TAKE
BIG	ERA	MAKE	THAN
BOON	ET	MAKING	THAT
BOTH	ETUDE	MANY	THE
BUT	EVIDENCE	MIT	THEIR
BY	EXAMPLE	MORE	THESE
CENT	FAR	MOST	THIS
CERTAIN	FILLS	MUCH	THROUGH
CHAPTER	FITS	MUST	THRU
CHEZ	FOLLOWING	NATURE	TO
COMING	FOR	NEEDED	TOO
COMMENT	FROM	NEW	TOWARD
COMMENTS	FULLY	NO	TOWARDS
COMPARISON	FUR	NOT	TYPE

stop list. As intermediate putput, all terms which are not in the stop list will be output in word frequency order. The editor will select those additional words which are to be stopped. These will be added to the stop list. The items will then be again passed against the expanded stop list. Each item will then be sorted by author and an index of all authors will be generated. The items will then be sorted on terms not in the stop list. A KWOC index will then be generated. Each item in each index will contain all authors, the title, and an appropriate bibliographical citation. Index terms which enrich the titles for retrieval purposes will be used as non stop list terms but will not be listed with the entry. Thus, one is likely to find a title such as IR on a Budget which may be enriched with a term such as "Information Retrieval" both under "Information" and under "Retrieval". An alternate sample KWOC index page is shown in Figure 4. Experiments are currently being conducted with various reduction ratios and formatting.

The system as proposed has all the advantages of speed and timeliness. It also has many of the deficiencies typical of such computer systems. We have attempted to reduce these by a moderate amount of human intervention.

SUMMARY AND CONCLUSION

The system which we propose for the literature of library and information science is a real system. It has been completely tested and is operational (for other interest fields) at the Technical Information Dissemination Bureau of the State

PARAMETERS
SEL. NO.-0106410 SOURCE-SOV PH SE R V- 1 P- 755 YR-67

PARTICLES
RADIOLYSIS OF DILUTE AERATED SULFURIC ACID SOLUTIONS WITH 5.3-MEV ALPHA PARTICLES
MARIANO MH SANTOS ML
SEL. NO.-0106403 SOURCE-RADIAT RES V- 32 P- 905 YR-67

FUEL MIGRATION INTO COATINGS OF PYROLYTIC-CARBON-COATED UO2 AND UC2 PARTICLES AS A RESULT OF IRRADIATION AND HEAT TREATMENT
MORGAN MT LONG EL HEWETTE DM
SEL. NO.-0106427 SOURCE-T AM NUCL S V- 10 P- 97 YR-67

FISSION-PRODUCT RELEASE DURING POST-IRRADIATION ANNEALING OF PYROLYTIC-CARBON-COATED FUEL PARTICLES -
MORGAN MT SISMAN O
SEL. NO.-0106428 SOURCE-T AM NUCL S V- 10 P- 98 YR-67

EXCITED PARTICLES AND THEIR ROLE IN ELEMENTARY PROCESSES OF RADIATION CHEMISTRY AND PLASMA CHEMISTRY. ((
TUNITSKI.NN KUPRIYAN.SE
SEL. NO.-0090112 SOURCE-AD-660408 V- AUG P- 60 YR-67

EFFECT OF IRRADIATION ON SUB-CELLULAR PARTICLES - DESTRUCTION OF SULPHYDRYL GROUPS
WILLS ED WILKINSO.AE
SEL. NO.-0106340 SOURCE-INT J RAD B V- 13 P- 45 YR-67

PARTICULATES
STUDIES OF EFFECT OF IRRADIATION OF CELLULAR PARTICULATES .6. COMPARISON OF UNCOUPLING AT 3 PHOSPHORYLATING SITES
YOST HT YOST MT ROBSON HH
SEL. NO.-0088912 SOURCE-BIOL BULL V- 133 P- 697 YR-67

PERFLUORODIMETHYLCYCLOHEX
RADIOLYSIS OF PERFLUORODIMETHYLCYCLOHEXANE .2. REACTOR RADIATIONS
SCOTT TH WETHINGT.JA
SEL. NO.-0106443 SOURCE-T AM NUCL S V- 10 P- 158 YR-67

RADIOLYSIS OF PERFLUORODIMETHYLCYCLOHEXANE .3. FISSION FRAGMENTS
SCOTT TH WETHINGT.JA
SEL. NO.-0106444 SOURCE-T AM NUCL S V- 10 P- 159 YR-67

PEROXY
REACTION RATES OF ALKYL AND PEROXY RADICALS WITH COPPER ION-PULSE RADIOLYSIS STUDIES
MACLACHL.A
SEL. NO.-0084189 SOURCE-J PHYS CHEM V- 71 P-4132 YR-67

PHENYLALANINE
RADIOLYSIS OF AQUEOUS SOLUTIONS OF PHENYLALANINE
BRODSKAY.GA SHARPATY.VA
SEL. NO.-0106405 SOURCE-RUSS J PH R V- 41 P- 583 YR-67

PHENYLAZO
PHOTOCHEMICAL REACTIONS OF AZO COMPOUNDS .14. PHOTOCHEMICAL CYCLODEHYDROGENATION OF 4,4$-BIS(PHENYLAZO)DIPHENYLMETHANE
JAMIESON NC LEWIS GE
SEL. NO.-0106312 SOURCE-AUST J CHEM V- 20 P-2777 YR-67

PHOSPHORESCENCE
OBSERVATIONS OF PHOSPHORESCENCE IN PLASTIC SCINTILLATORS EXCITED BY 60CO GAMMA IRRADIATION
KROPP JL
SEL. NO.-0084181 SOURCE-J CHEM PHYS V- 47 P-3340 YR-67

PHOSPHORYLATING
STUDIES OF EFFECT OF IRRADIATION OF CELLULAR PARTICULATES .6. COMPARISON OF UNCOUPLING AT 3 PHOSPHORYLATING SITES
YOST HT YOST MT ROBSON HH
SEL. NO.-0088912 SOURCE-BIOL BULL V- 133 P- 697 YR-67

PHOTOCHEMICAL
A PHOTOCHEMICAL STUDY OF SPIRO-PYRAN ADSORBED ON SILICIUM GEL
BALNY C DJAPARID.K DOUZOU P
SEL. NO.-0106326 SOURCE-CR AC SCI C V- 265 P-1148 YR-67

PHOTOCHEMICAL ISOMERIZATIONS OF SOME NATURAL PRODUCTS AND RELATED COMPOUNDS
CROWLEY KJ
SEL. NO.-0106304 SOURCE-ACT CIENT V V-S 18 P- 318 YR-67

TRANSIENT SPECIES PRODUCED IN PHOTOCHEMICAL DECOMPOSITION OF CERIC SALTS IN AQUEOUS SOLUTION . REACTIVITY OF NO3 AND HSO4 FREE RADICALS
DOGLIOTT.L HAYON E
SEL. NO.-0084184 SOURCE-J PHYS CHEM V- 71 P-3802 YR-67

PHOTOCHEMICAL SYNTHESIS OF ALKYL- AND ARYL-CYCLOPROPENES
DURR H
SEL. NO.-0106310 SOURCE-ANGEW CHEM V- 6 P-1084 YR-67

PHOTOCHEMICAL TRANSFORMATIONS.15. PHOTOCHEMICAL ISOMERIZATION OF EXO-TRICYCLO(4,2,1,02,5)NONA-3,7-DIENE SYSTEM
EBERBACH W PRINZBAC.H
SEL. NO.-0088922 SOURCE-HELV CHIM A V- 50 P-2490 YR-67

PHOTOCHEMICAL CONVERSION OF COLUPULONE INTO 4-DEOXYCOHUMULONE
FERNANDE.CM
SEL. NO.-0084174 SOURCE-CHEM COMM V- P-1212 YR-67

PHOTOCHEMICAL REACTION OF TECHNETIUM(4) CHLOROCOMPLEXES
FUJINAGA T KOYAMA H KANCHIKU.Y

Figure 4 Sample KWOC Index

University of New York at Buffalo. Although we believe that a current awareness service for library and information science is needed and, as we have pointed out in the Introduction, many authors have reinforced our belief, we feel it would be desirable to test the hypothesis that current awareness for our field is needed and wanted. For this reason, we are preparing a proposal (to be completed by July, 1968) for a pilot current awareness service for our field. This proposal will be submitted to appropriate agencies for funding.

BIBLIOGRAPHY

1. A.L.A. Bulletin, v. 61, no. 8, Sept. 1967, p. 951-954.
2. Library Trends, v. 15, nos. 3 and 4, Jan., Apr. 1967.
3. Whatley, H. Allan, "Library Science and Documentation Indexing and Abstracting Services," Unesco Bulletin for Libraries, v. 19, no. 5, Sept./Oct. 1965, p. 226-231.
4. Gilchrist, Alan, "Documentation of documentation," Aslib Proceedings, v. 18, no. 3, March, 1966, p. 62-80.
5. Whatley, H. Allan, A Survey of the Major Indexing and Abstracting Services for Library Science and Documentation, London, The Library Association, 1966.
6. Richmond, Phyllis A., "Systems Evaluation by Comparison Testing," College and Research Libraries, v. 27, no. 1, Jan. 1966, p. 30.
7. Whatley, op. cit, 1966, p. 62.
8. System Development Corporation, National Document Handling Systems for Science and Technology, New York, Wiley, 1967, Chapter Four, "Document Users," p. 102-112.

9. Herner, Saul and Herner, Mary, "Information Needs and Users in Science and Technology," (In Annual Review of Information Science and Technology, v. 2, 1967, New York, Wiley, 1967, p. 1-34.

10. Tukey, J. W. "The Citation Index and the Information Problem: opportunities and Research in Progress," Annual Report, Princeton University, Princeton, N. J., 1962.

11. Cleverdon, C. W. "Automation in Indexing," ASLIB Proceedings, v. 13, p. 107-109 (1961).

12. Janaske, P. C. "Manual Preparation of a Permuted-Title Index," BSCP Communique 7-62, Biological Abstracts, Phila., Pa., June 1962.

13. Skaggs, B. and M. Spangler, "Easing the Route to Retrieval with Permuted Indexes", Business Automation, v. 9, p. 26-29, 60, (1963).

14. Kennedy, R. A. "Library Applications of Permutation Indexing," J. Chem. Documentation 2, 181-185 (1962).

15. Stevens, Mary Elizabeth, "Automatic Indexing: A State-of-the-Art Report," NBS Monograph 91, March 30, 1965, p. 56.

16. Ruhl, M. J. "Chemical Documents and Their Titles: Human Concept Indexing vs. KWIC - Machine Indexing", unpublished paper presented at the 144th National Meeting, ACS, Los Angeles, Cal. 2 Apr. 1963, 12 p.

17. Stevens, op. cit, p. 59.

18. Lane, B. B. "Key Words In--and Out of--Context," Special Libraries, 55, 45-46 (1964).

19. Montgomery, C. and D. R. Swanson, "Machine Like Indexing By People", Amer. Documentation, v 13, p. 359-366, (1962).

20. National Science Foundation, Current Research and Development in Scientific Documentation, No. 11 (NSF-63-5) Nov. 1962, 440 p. U. S. Government Printing Office, Washington, D. C.

21. Cleverdon, C. W., "The Evaluation of Systems Used in Information Retrieval," in International Conference on Scientific Information, Proceedings 1959, v. 1, p. 687-698.

22. Black, D. V., "Indexing Techniques Description and Background," Appendix to "Document Storage and Retrieval Techniques," Planning Research Corp., Los Angeles, 13, June 1963.

23. Stevens, op. cit, p. 160.

24. Black, J. D., "The Keyword: Its Use in Abstracting, Indexing and Retrieving Information," Aslib Proceedings, v. 14, p. 313-321, (1962).

25. Wyllys, R. E., "Automatic Analysis of the Contents of Documents Part I: Historical Review," Field Note FN 6089, System Development Corp., Santa Monica, California, 1961.

26. Menzel, Herbert, "The Information Needs of Current Scientific Research," (In Chicago University. Graduate Library School, Library Catalogs, Changing Dimensions, Chicago, University of Chicago Press, 1964). Published originally in Library Quarterly, Jan. 1964.